

Declaration

I, Xi Yongzhi, hereby declare as follows:

The undersigned declares further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of this application or any patent issuing thereon.

I am the inventor and applicant of the US Patent Application Serial No. 10/534,538.

Although we submitted the polypeptide sequences (Genbank accession number AAK98621) on 19 July 2001 with the US NCBI-Genbank, however, it had never been published on or after that day because we withdraw it immediately, the information respect the polypeptide was empty until we updated the new information on Feb 8, 2006. The detailed explanations are as follows:

I. Observation on relevant date of the sequences in the present patent.

As we know, US NCBI-GenBank, established by the National Center for Biotechnology Information of National Institutes of Health, collects and arranges numerous nucleotide sequences of genes and the corresponding amino acid sequences provided by biological scientists from every country in the world.

Here, we would like to remind the US examiner that: the date, mentioned in the LOCUS line recorded in the NCBI-GenBank data, is the date when the data is finally published. In most of cases, it is also the date when the data is first published. Another recorded date (see below) is the date when the sequence is submitted with the databank. It should be noticed that the two dates are not guaranteed under laws and the databank does not declare that the two dates are correct. *(the relevant*

original text: "The date on the locus line is the date the record was last made public. If the record has not been updated since being made public, the date would be the date that it was first made public. If any of the features or annotations were updated and the record was rereleased, then the date corresponds to the last date the entry was released. Another date contained in the record is the date the record was submitted (see below) to the database. It should be noted that none of these dates is legally binding on the promulgating organization. The databases make no claim that the dates are error -free." see Ilene Karsch-Mizrachi and B. F. Francis Ouellette. THE GENBANK SEQUENCE DATABASE. Chapter 3, pp51, last paragraph; in *Bioinformatics: A practical guide to the analysis of gene and proteins. Second Edition.* Andreas D. Baxevanis & B.F.Francis Ouellette., eds. Copyright©2001 John Wiley & Sons, Inc. ISBNs: 0- 471-38390-2 (Hardback) : 0-471-38391-0 (Paper): 0-471-22392-1 (Electronic)). Hence, the time or date for the sequences recorded in NCBI-GenBank is just provided to the reader for reference only, and cannot be used as the law grounds for arbitration.

II. Observation on the inventiveness of the sequence.

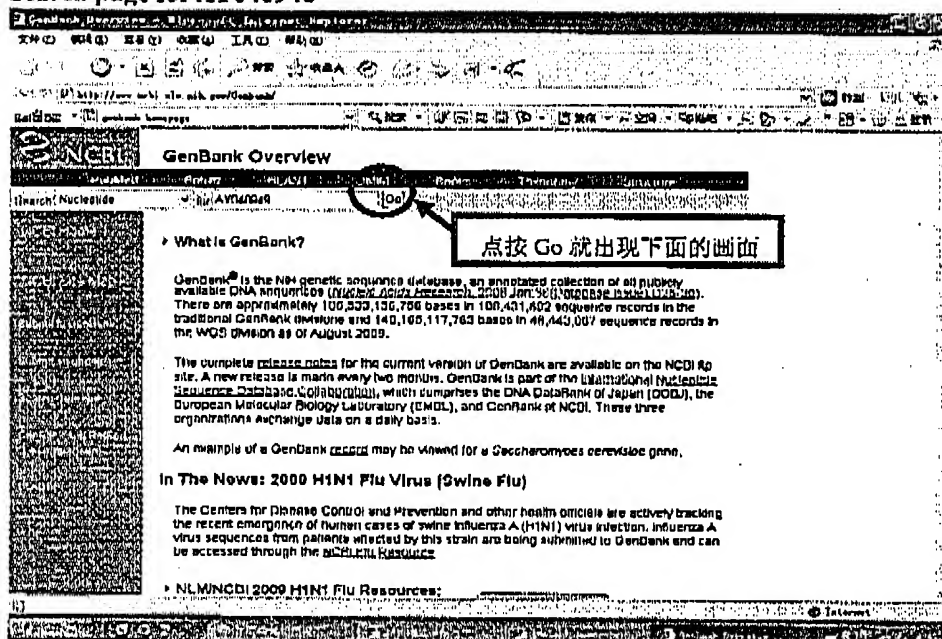
The accession number AAK98621 refers to the amino acid sequence of the chicken type II collagen which is translated by the mRNA (AY046949) of the chicken type II collagen. As we declared in June 29, 2009, AAK98621 and AY046949 indeed were submitted with US NCBI-GenBank on July 19, 2001. However, taken into multiple factors, we withdrew the submission of the two sequences from NCBI-Genbank immediately, and also withdrew the records on relevant sequences in NCBI-Genbank. Only the corresponding accession number and filing date are left. More importantly, NCBI-Genbank does not disclose these sequences (see NCBI-GenBank search page and detailed record Attachments 1 and 2). That is, AAK98621 is formally published on February 8, 2006, much latter than 2002 when the patent was filed. Hence,

the novelty and the inventiveness of the present patent is not taken away or interfered.

The publication date of AAK98621 and AY046949 are further checked as follows. The search results of all the history records of AAK98621 and AY046949, are separately summarized based on author and accession number:

1. AY046949, version AY046949.1 GI:15546069, is the version of the mRNA of chicken type II collagen submitted on February 8, 2006, with a fragment of 4837bp in length. Although it is shown that version AY046949 is first published on July 19, 2001, no relevant nucleotide sequence can be found in NCBI-GenBank. Namely, the sequence submitted on July 19, 2001 is not published. Its novelty and inventiveness are preserved. AY046949 is formally published on February 8, 2006(see NCBI-GenBank search page below and detailed record attachment 1).

Search page for AY046946



NCBI Nucleotide

Search Nucleotide:

Format: GenBank EMBL DDBJ

Galus gallus alpha 1 type II collagen precursor (COL2A1) mRNA, complete cds, alternatively spliced

Features: **Sequence:**

LOCUS AT040940 4837 bp mRNA linear VET 06-YES-3006

DEFINITION Gallus gallus alpha 1 type II collagen precursor (COL2A1) mRNA, complete cds, alternatively spliced.

ACCESSION AT040940

VERSION AT040940.1 GI:15546070

KEYWORDS

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
Eubacteria; Bacteria; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Sphenoceras; Psittaciformes; Theropoda; Coeloceras; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
[J04001 to 4837]

REFERENCE

AUTHOR Mi, G., Liu, M., Liang, F., Qiu, S., Sun, Y., Yang, F. and Li, Y.

TITLE Recloning, characterization and localization of chicken type II procollagen gene

JOURNAL Gene 366 (1): 57-65 (2006)

PMID 16297552

Sequence Analysis Tools

- ▶ BLAST Sequence
- ▶ Pick Primers

Articles about the COL2A1 gene

- ▶ Molecular cloning, characterization and localization of chicken type II procollagen (COL2A1) gene and sequence of the chicken type II procollagen gene Char U. Gen. 1994; 150(1): 1-10
- ▶ Identification of genomic DNA coding for chicken type II procollagen (COL2A1) gene

Related mRNA

See the reference mRNA sequence for the COL2A1 gene (NM_004420.1)

2. AAK98621, version AAK98621.1 GI:15546070, is the version of the amino acid sequence of chicken type II collagen submitted on February 8, 2006, with a fragment of 1420aa in length. Although it is shown that version AAK98621 is first published on July 19, 2001, no relevant amino acid sequence can be found in NCBI-GenBank. Namely, the sequence submitted on July 19, 2001 is not published. Its novelty and inventiveness are preserved. AAK98621 is formally published on February 8, 2006 (see NCBI-GenBank search page below and detailed record attachment 2).

Search page for AAK98621

GenBank Overview

What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. (NCBI's GenBank accession numbers range from 1 to 100,000,000). There are approximately 100,000,000 sequences in the GenBank database and 100,000,000 records in the traditional GenBank divisions and 100,000,000 records in the WGS division as of August 2009.

The complete release notes for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA Database of Japan (DDJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

An example of a GenBank record may be viewed for a Saccharomyces cerevisiae gene.

In The News: 2009 H1N1 Flu Virus (Swine Flu)

The Centers for Disease Control and Prevention and other health officials are actively tracking the recent emergence of human cases of swine influenza A (H1N1) virus infection. Influenza A virus sequences from patients infected by this strain are being submitted to GenBank and can be accessed through the NCBI Flu Resource.

NCMNCBI 2009 H1N1 Flu Resources

Protein

Alpha 1 type IIA collagen precursor (Gallus gallus)

Try the [Sequence Viewer](#) for a more informative view of the biological features.

Sequence Analysis Tools

- BLAST Sequence
- Conserved Domains

Articles about the COL2A1 gene

- Molecular cloning, characterization and localization of chicken type II pro-collagen gene.
- Structure and sequence of the chicken type II procollagen gene. Chai J, Liu Y, Yang Y, et al. (1994)
- Identification of genomic DNA coding for chicken type II procollagen. Chai J, Liu Y, Yang Y, et al. (1994)

Identical Protein for AA05021.1

- collagen, type II, alpha 1 (Gallus gallus)

NCBI

3. AF452711 is the genomic sequence of chicken type II collagen. Version AF452711.2 GI: 31340541, is the newest version of the genomic sequence of chicken type II collagen submitted on February 8, 2006. The fragment comprises exons 8-52 and has 12003bp in length. However, version AF452711.1, updated on June 3, 2003, is part of the genomic sequence of chicken type II collagen, with only 5494bp in length. The two are different by 7509 bp. Although it is also shown that version AF452711 was first submitted on November 27, 2001, just like AY046949 and AAK98621 sequence, no corresponding genomic sequence can be found in GenBank (see NCBI-GenBank search page below and detailed record attachments 3 and 4) .

Search page for AF452711

GenBank Overview

What's GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (<http://www.ncbi.nlm.nih.gov/genbank/>). There are approximately 100,000,000,000 bases in 100,000,000,000 sequence records in the traditional GenBank divisions and 140,100,117,763 bases in 40,440,007 sequence records in the WGS division as of August 2009.

The complete (CCO) database for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

An example of a GenBank record may be viewed for a *Saccharomyces cerevisiae* gene.

In The News: 2009 H1N1 Flu Virus (Swine Flu)

The Centers for Disease Control and Prevention and other health officials are actively tracking the recent emergence of human cases of swine influenza A (H1N1) virus infection. Influenza A virus sequences from patients affected by this strain are being submitted to GenBank and can be accessed through the [NCBI Flu Resources](http://www.ncbi.nlm.nih.gov/genbank/).

• NLM/NCBI 2009 H1N1 Flu Resources

• Newest 2009 H1N1 Influenza A sequences

7

Attachment 1: Search record for AY046949, version AY046949.1 GI:15546069

LOCUS AY046949 4837 bp mRNA linear VRT 08-FEB-2006

DEFINITION *Gallus gallus* alpha 1 type IIA collagen precursor (COL2A1) mRNA, complete cds, alternatively spliced.

ACCESSION AY046949

VERSION AY046949.1 GI:15546069

KEYWORDS

SOURCE *Gallus gallus* (chicken)

ORGANISM *Gallus gallus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 4837)

AUTHORS Xi, C., Liu, N., Liang, F., Guo, S., Sun, Y., Yang, F. and Xi, Y.

TITLE Molecular cloning, characterization and localization of chicken type II procollagen gene

JOURNAL *Geno* 366 (1), 67-76 (2006)

PUBMED 16297573

REFERENCE 2 (bases 1 to 4837)

AUTHORS Caixia, X., Yongzhi, X., Siqu, G. and Yiyang, S.

TITLE Direct Submission

JOURNAL Submitted (19-JUL-2001) Immunology, Beitapinglu Hospital, Beitapinglu, Beijing 100039, China

FEATURES

source Location/Qualifiers

1..4837

/organism="Gallus gallus"

/mol_type="mRNA"

/db_xref="taxon:9031"

/tissue_type="sternal"

/dev_stage="17-day embryo"

/country="China"

/note="breed: San-Huang"

gene 1..4837

/gene="COL2A1"

CDS 15..1307

/gene="COL2A1"

/note="lacks exon 2; alternatively spliced"

/codon_start=1

/product="alpha 1 type IIA collagen precursor"

/protein_id="AAK98621.1"

/db_xref="GI:15546070"

/translation="MIQRRPPRSAAALLLLLLLTAAAAQDRDLRQPGPKGQKQEPGD
IKDVVGPRGPPGPGPAGEQGGQGRGDRGEKGEKAPGPRGRDGEPTGPNPGPPGPPG
PGPPGLGGNFAAQMAGGFDEKAGGAQMGVMQGMGPMGPRGPPGPTGAPGPQGFQGNP

批注: The serial number of the nucleic acid sequence

批注: the sequence of other vertebrate

批注: This is the time when the sequence was submitted for the second time, and also is the time when the sequence was first published.

批注:

批注:

批注: The sequence was submitted for the second time after the paper was published.

批注:

批注:

批注: This is the date when the sequence was first submitted before the withdrawal from the Genbank; but the nucleotide sequence submitted this time was not disclosed, no relevant sequence can be found in Genbank.

批注: AAK98621 (the serial number of the amino acid sequence corresponding to AY046949)

GEFGEPGAAGPMGPRGPPCPGKPGDDGETGKPGKSGERGPPCPQGARGFPGTGLPG
 VKGHRGYPGLDGAKGFAAGAPGAKGSSGSPGSPGMPGRCIPGERGRPGPSGAAGA
 RGNGLPCPAGPPGPGVAGAPGFPAGPSKGEAGPTGARGPBGAGPGRGESGTPGSP
 GPAGAPGNPGTDGIPGAKGSAGAPGIAGAPGFPGRGPPGPGQATGFLGPKGQTGEPC
 IAGFKGEQGPKEGTGPAGPQCAPGACGEGKRGARGEPGAAGPVGPPGERCAPGNRGF
 PGQDGLACPKGAPGERGPAAGLAGPKGATGDCRPGEPGLPGARGLTGRPDAGPQGV
 GFTGAPGEGDRPGPPGPGQARGQPGVMGFPGPKGANGEPGKAGEKGLPGAPOLRGLPG
 KDGETGAAGIPGPAGPVGERGEQCAPGSPCFQGLPGPPGPPGSGKPGDQGVPGEAGA
 PGLVGPGRGERGPPGERGSPGAQGLQGPRGLPGTPTGDPKGATGPAGPNAQGPPLQ
 GMPGERGAAGIAGLKGDGVDGEGKPGFAGAPCKDGARCLTGPICPPGAPNGEKGESG
 PPGPSGAACARGAPGERGEPGAPGAGFAGPPGADGQPGAKGEQCEPCQKCDAGAPGP
 QGPGSAGPQPQPTGVTGPKGARGAQGPPGATGPPGAAGRVGPPGPNNGPPGPPGSA
 GKDGPKGVRCDAQPPCRAGDPGLQGPAIPGEGKEPGEDEGAPGDPGPPGQGLAGQRC
 LVGLPGQGRGERGPPGLPGSPGEPKQCAPGSAGDRGPPGPPGPPGLTGAPGEPGREGN
 PGADGLPCRDGAAGVKGDGRTGFPVAGAPGAPGAPGPPVGTGKQGDGRTGAGQGM
 GPSGPAGARGMPGPPGPPGPRCKGCTGEGAGERGLKGRGFTGLQLPGPPGPPGSDQGAAG
 PAGPSGPPRGPVCPGSKDGSNGMPGPIGPPGPPGRSGEPGAPPPGNPPGPPGPPG
 PCTGIDMSAFAGLGQTEKGPDPPIRYMRADAAAGGLRQINDVEVDATLKSNNQIESIRS
 PEGSKKNPARTCRDIKLCHPEWKSQDYWIDPNQGGTLDALKVFCNMETGETCVYPTPS
 SIPRKNWWTSTKTKDKKHVWFPAETINGGPHFSYGDENLSPNTASIQMTFLRLSTEGSQ
 NVTYHCKNSIAYMDEETGNLKAITLQGSNDVEIRAEGNSRFTYSVLEDGCTKHTGKW
 GKTVTIEYRLQKTSRLSTVDTAPMDIGGADQEPGVLDIGPVCFLL

sig_peptide 45..119
 /gene="COL2A1"
misc_feature 120..386
 /gene="COL2A1"
 /note="Region: N-propeptide"
misc_feature 387..461
 /gene="COL2A1"
 /note="Region: N-telopeptide"
misc_feature 462..3485
 /gene="COL2A1"
 /note="Region: triple helical domain"
misc_feature 3486..3566
 /gene="COL2A1"
 /note="Region: C-telopeptide"
misc_feature 3567..4304
 /gene="COL2A1"
 /note="Region: C-propeptide"

ORIGIN

1 ucgcctcgcc ccgacgggacg gtagaggggac ctcgcgcgcg agcagtgaga ggccgcgcgc
 61 cggcccgctc cgcgcctctc ctctctcttc tctctcttct ctcggccgac gcagcccgcc
 121 agggccgcga cctccgacaa cctggcccca agggacagaa gggagagccc ggagatatta
 181 aagatgttgt aggaccccca gggcctcagc ggcgcgcggc ccgcgcgcga ggcacgggac

1022: Nucleotide sequence
 (mRNA)

241 agcagagagga caglygagag aagagagagga aggylgctcc tggccccagt ggglyggatg
301 agagagagag cagccctaga aagccagggc ccccgagctc ccccgagctc cctgagagag
361 cagagagagag aggagagag cagagagagag tggagagagag cttcagagag aggagagag
421 agagagagag aggagagag cagagagagag tggagagagag cagagagagag aggagagag
481 ccccgagagag aggagagag ccccgagagag ttcagagagag ccccgagagag ccccgagagag
541 ccccgagagag tggagagag aggagagagag aggagagagag aggagagagag aggagagagag
601 agagagagag agagagagag ccccgagagag cttcagagag tggagagagag aggagagagag
661 agagagagag cttcagagag aggagagagag tggagagagag aggagagagag aggagagagag
721 ccccgagagag tggagagagag aggagagagag aggagagagag aggagagagag aggagagagag
781 ccccgagagag aggagagagag ccccgagagag tggagagagag aggagagagag aggagagagag
841 aggagagagag ccccgagagag aggagagagag cttcagagag tggagagagag aggagagagag
901 ccccgagagag tggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
961 agagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1021 agagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1081 aggagagagag cttcagagag aggagagagag tggagagagag aggagagagag aggagagagag
1141 ccccgagagag aggagagagag ccccgagagag aggagagagag aggagagagag aggagagagag
1201 agagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1261 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1321 ccccgagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1381 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1441 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1501 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1561 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1621 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1681 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1741 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1801 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1861 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1921 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
1981 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2041 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2101 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2161 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2221 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2281 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2341 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2401 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2461 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2521 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2581 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2641 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2701 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2761 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag
2821 aggagagagag aggagagagag aggagagagag cttcagagag aggagagagag aggagagagag

2881 cgggtagccg aggtccccc agccccgtgg gccccttga actgacgggl cctgctggag
2941 aacccgrrcr crugggcnnc cctgggtcra acggtctccc ngyccgggat ggcgcaacra
3001 acrtgagggg lgtatgttgt raaacccrcc ctgtgggagc ccccgrract cctggagccc
3061 ctggcgcgcc cggccctgtt ggtcccacra raaacaagg agnccggggc gggcrrrrrr
3121 cacaaggccc calgggagcc tctggtcccg ctggagctcg agggatgccc rrtccccaag
3181 gacctcrrrr taaacaaggg gggcggggg aggtcrrrra raaagggtcg nngggccccc
3241 gtgggllcnc cgtctgrrrr rrtctgccc gacccccgg cccgtctgga gaccagrrrr
3301 ctgcccgttc cgtgrrrrcc lccgggtccc gaggcccccc tgggtccgrr rrrccctctg
3361 gcaaaarrrr cttcaacrrc atgacggccc ccaggggccc tcccgggtcc crrrracrra
3421 gtgggtgccc cggccctgrr rrtctcctrr raaacccgg tctcgggtt cclccggccc
3481 ccccgrrrac cggcatagac atgtatgall lggclggact ggtcagrrrr rrrrrrrrrc
3541 ccgacccncl ccgclacalg agggcagrrr rrrcrrrrrr rrrrrrrrrr cagcagrrrr
3601 lggrrrrrrr tggccacctc aaatccctca ncnatcrrr tgggagrrrr cggcggcccc
3661 aggggtccrr gggggacccr gcccrrrrrr rrrrrrrrrr caaacrrrrr crrccrrrrr
3721 rraaagrrrr arattactgg attgacccrr accnngggcg cnccttggrr ggcrrrrrrr
3781 ttttclgrrr calgggagrr ggcrrrrrrr ggcrrrrrrr gccccccrrr agcrrrrrrr
3841 ggaagrrrrr gtggacrrrr agnccrrrrr ncnrrrrrrr cgtctggrrr gcnrrrrrrr
3901 tcacrrrrrr tttccacttc arctacrrrr atgrrrrrrr rrrcccccrr accrrrrrrr
3961 tccgrrrrrr clccclgrrr clccclgrrr ccggggrrrr ccggggrrrr accrrrrrrr
4021 rcaagrrrrr crrrrrrrrr atgrrrrrrr rrrrrrrrrr cctgrrrrrr rrrrrrrrrr
4081 tccrrrrrrr crrrrrrrrr rrrrrrrrrr ccggggrrrr ccggggrrrr accrrrrrrr
4141 tcllggrrrr cggclgrrrr ncnrrrrrrr gcnrrrrrrr crrrrrrrrr atcrrrrrrr
4201 rrrrrrrrrr rrrrrrrrrr crrrrrrrrr tagrrrrrrr accrrrrrrr attggrrrrr
4261 crrrrrrrrr rrrrrrrrrr rrrrrrrrrr crrrrrrrrr crrrrrrrrr grrrrrrrrr
4321 tlllrrrrrr llllrrrrrr grrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr
4381 grrrrrrrrr grrrrrrrrr crrrrrrrrr aaagrrrrrr rrrrrrrrrr crrrrrrrrr
4441 crrrrrrrrr rrrrrrrrrr rrrrrrrrrr grrrrrrrrr rrrrrrrrrr crrrrrrrrr
4501 crrrrrrrrr crrrrrrrrr rrrrrrrrrr aaagrrrrrr rrrrrrrrrr rrrrrrrrrr
4561 grrrrrrrrr rrrrrrrrrr rrrrrrrrrr grrrrrrrrr grrrrrrrrr crrrrrrrrr
4621 grrrrrrrrr crrrrrrrrr rrrrrrrrrr rrrrrrrrrr grrrrrrrrr rrrrrrrrrr
4681 crrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr
4741 rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr
4801 tgggtggrrr grrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr

Attachment 2: Search record for AAK98621, version AAK98621.1 GI:15546070

LOCUS AAK98621 1420 aa linear VRT 08-FEB-2006
 DEFINITION alpha 1 type IIA collagen precursor [Gallus gallus].
 ACCESSION AAK98621
 VERSION AAK98621.1 GI:15546070
 DBSOURCE accession AY046949.1
 KEYWORDS .
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (residues 1 to 1420)
 AUTHORS Xi, C., Liu, N., Liang, F., Guo, S., Sun, Y., Yang, F. and Xi, Y.
 TITLE Molecular cloning, characterization and localization of chicken
 type II procollagen gene
 JOURNAL Gene 366 (1), 67-76 (2006)
 PUBMED 16297573
 REFERENCE 2 (residues 1 to 1420)
 AUTHORS Caixia, X., Yongzhi, X., Siqi, G. and Yiyang, S.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUL-2001) Immunology, Beitapinglu Hospital,
 Beitapinglu, Beijing 100039, China
 COMMENT Method: conceptual translation supplied by author.
 FEATURES
 Location/Qualifiers
 source 1..1420
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="sternal"
 /dev_stage="17-day embryo"
 /country="China"
 /note="breed: San-Huang"
 Protein 1..1420
 /product="alpha 1 type IIA collagen precursor"
 sig_peptide 1..25
 Region 26..114
 /region_name="N-propeptide"
 Region 115..139
 /region_name="N-telopeptide"
 Region 140..1147
 /region_name="triple helical domain"
 Region 1148..1174
 /region_name="C-telopeptide"

批注: The serial number of the amino acid sequence

批注: the sequence of other vertebrate

批注: This is the time when the sequence was submitted for the second time, and also is the time when the sequence was first published.

批注: Corresponding mRNA serial number, the first filing date and the amending date of the sequence are completely consistent with those of AAK98621 (its corresponding amino acid sequence).

批注: The sequence was submitted for the second time after the paper was published.

批注: This is the date when the sequence was first submitted before the withdrawal from the Genbank; but the amino acid sequence submitted this time was not disclosed, no relevant sequence can be found in Genbank.

Region 1175..1420
 /region_name="C-propeptide"
Region 1185..1420
 /region_name="COLFI"
 /note="Fibrillar collagens C-terminal domain: Found at C-termini of fibrillar collagens: Ephedratia muelleri procollagen EMF1alpha, vertebrate collagens alpha(1)III, alpha(1)II, alpha(2)V etc; smart00038"
 /db_xref="CDD:47387"
CDS 1..1420
 /gene="COL2A1"
 /coded_by="AY046949.1:45..4307"
 /note="lacks exon 2; alternatively spliced"

ORIGIN |

```

1  mhgrppprsa allllllllt aaaaaqdrdl rqpqpkqkqg cpqdikdvvg prppppqzpr
61  agcagqrgdr gokgokgng prgrdgerkt pknppppppp spppppglpg nfnanmggff
121  dekaggaqmg vmqgpmgpmg prpppgntga pgnqgfgnnp kepkepkagk pmkprgppgp
181  pggpdddel gkpkksgrg ppqpqargf pgtnglpgvk ghrgypgldg akgaagapga
241  kkesgspgen gspgmgprg lpgergprp sgaagargnd rldgpagppg pvgpagaapgi
301  pgapgskega gntgarfpeg uqgrkesgt pksppagap gnpgtgipg akgsagapgi
361  agapgfppr gpppqqgatg plgpkgtgce pgiagfkgoq gpkgotgpag pggapspage
421  egkrargrge kaaapvppp erkapnrgf pggdglagpk gapgerpag lagpkgatgd
481  pgrpgepglp gargltgrpg dagpqqkvpp tgapgodgrp gppgnqarg qpgvmgfnep
541  kgangepkka gekglpkkk lrglpkkgc tgaagppga gpvgergcqg apgspgfqgl
601  pgnppppres gkpgdqrvp egaapglvp rgerkfper aspaaqlqg prglpgrpt
661  dgpkgatgpa gpnagaqppp lqmpgorga agiaglkgrd gdvgekgpeg apgkdgarcl
721  tgpigppgpa gpngekgess pnpssnaga rgapgorp gnpagpagn ppgadgppga
781  kgcqgcpqk gdagapppg psapppqgp tgvlsppkar gaqgppgatg lpgaagrvgp
841  pppnppppp ppppsagkdg pkgvrgdagn pgragdpqlg gpagppgokg opgodepagp
901  dgppppqela gqrgivglpg qrgergfpl ppspgepkq guppsuadrn pppnppppp
961  kppagepge gnpadglpg rdgaagvkgd rgotgppvap gnpagpagn pvgptgkqgd
1021  rgetgaqpm ppsppagarg mpqpqprgd kgetgoagor glkghrgftg lqglpappgp
1081  sgdaqaaaga ppsppppp ppsppkdkk ngmpgplgpp gprgrsgpge pagppppppp
1141  pppppppgtg idmsafaglg qtekspdir ymradeaagg lrqhdvevda tlkslnnqle
1201  slrspegskk nparterdik lchpewksd ywidpqqct ldnikvfcm otgetcvypt
1261  pssiprknw tskldkdkhv wfuetinggf hfsygdnl pntasiqmtf lrlstegsq
1321  nvtyhcknsi aymdeotgnl kkailiqsn dveiraogns rftysvledg ctkhtgkwk
1381  tvioyrlqkt srslivdtap mdlggadqef gvdigpvcfl

```

批注: The amino acid
 sequence

Attachment 3: Search record for AF452711, version AF452711.1 GI:28190290

LOCUS AF452711 5494 bp DNA linear VRI 01-FEB-2003
 DEFINITION Gallus gallus alpha 1 type II procollagen (COL2A1) gene, partial cds.
 ACCESSION AF452711
 VERSION AF452711.1 GT:28190290
 KEYWORDS
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 5494)
 AUTHORS Caixia, X., Yongzhi, X., Siqi, G. and Yuying, S.
 TITLE Gallus gallus alpha 1 type IIA collagen precursor (COL2A1)
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 5494)
 AUTHORS Caixia, X. and Yongzhi, X.
 TITLE Direct Submission
 JOURNAL Submitted (27-NOV-2001) Immunology, Beitaipinglu Hospital,
 Beitaipinglu, Beijing 100039, China
 COMMENT [WARNING] On Jun 3, 2003 this sequence was replaced by GI:31340541.
 FEATURES
 Location/Qualifiers
 source 1..5494
 /organism="Gallus gallus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9031"
 gene <1..5494
 /gene="COL2A1"
 mRNA join(<1..21,103..156,239..292,387..440,530..637,
 746..799,913..966,1235..1396,1489..1596,1906..2013,
 2174..2227,2325..2432,2701..2754,2864..2971,3106..3159,
 3328..3435,3515..3803,4211..4398,4993..5235,
 5348..5494)
 /gene="COL2A1"
 /product="alpha 1 type II procollagen"
 CDS join(<1..21,103..156,239..292,387..440,530..637,
 746..799,913..966,1235..1396,1489..1596,1906..2013,
 2174..2227,2325..2432,2701..2754,2864..2971,3106..3159,
 3328..3435,3515..3803,4211..4398,4993..5235,5348..5494)
 /gene="COL2A1"
 /codon_start=1
 /product="alpha 1 type II procollagen"

批注: The nucleotide serial number of the genome

批注: the sequence of other vertebrate

批注: the time when it was submitted for the second time

批注: This is the date when the sequence was first submitted before the withdrawal from the Genebank; but the nucleotide sequence submitted this time was not disclosed, no relevant sequence can be found in Genebank.

批注: It was replaced by sequence GI:31340541 on June 3, 2003.

```

/protein_id="AA033039_1"
/db_xref="GI:28190291"
/translation="PGKDGARGLTGPTGPPGAPNGEKGESGPPPSGAAGARGAPG
ERGEIPGAPGPAGFAGPPGADGQPGAKGEQGEPCQKGDAGAPGPQGPSGAPCPQGPVGV
TGPKGARGAQGPCCATGTPGAAGRVPGPNGNPGPPGPGSAGKDGPKGYRGDAGPP
GRAGDPGLQGPAGPPCEKCEPGEUGPAGPDGPPGPQGLACQRGIVGLPGQGRGGRFP
LPGPSGEPCKQCAPGSAGDRGPPGPVGPCLTGPAGEPGRGNPGADGPPGRDGAAGV
KGDREGTGPVGAAPGAPGAPGPVGPCTCKQGDREGTGAQGPMPGSPAGARGMPGPQ
GPRGDKGETGEACERGLKCHRGFTGLQGI.PGPPGPSGDQCAAGTAGPSGPRGPPGPVG
PSCKDGSNGMPGPIGPPGPRGRSGEPGAPGPPGNPGPPGPPGPGTCTIDMSAFAGLQ
TEKCPDPIRYMGADEAAGGLRQHDVEVDATLKSNNQIESIRSPGSKKNPARTCRDI
KLCHPEWKSQDYWDPNQCCTLDALKVFCNMETGETCVYPTSSIPKKNWTSKTKDK
KIVWFARTINGGFHFSYGDENLSPNTASTQMTPLRLSLTEGSQNVYIICKNSIAYMDE
ETGNLKKATLTCQSNDEIRAGNSRFTYSVLEDCCTKHTGKWKTVIEYRSQKTSRI
PTVDIAPMDIGGADQEFQVDTGPVCFL"

```

批注: The serial number of the corresponding amino acid sequence, the first filing date and the amending date of the sequence are completely consistent with those of AF452711 (its corresponding nucleotide sequence).

ORIGIN

```

1  ccaggcaagg atggcgccacg ttaagtgagg gacggcccat aagataggct gggaaagcat
61  gclccacagag accacataccl cclctctctc tctctcccat agggctctgc gggctccatt
121  aatcccccctg gccclycttg ccccaacagt gggaaagcat gacgacatc acngcccc
181  acatllacgcc ccataaagtg accccatgic ctccacctct cctcccttcc tttccagag
241  tgaatccggc cctcclyctc catctaatgc tgcgggtgac cgtggtgccc ccgtangenc
301  aatgtctgaa gcccctggct gccctaaccc ttcuuccctaa accccctcca accccttcl
361  amccctcccc catctcttcc cctllgggag agcctgagca acccggtgac cccggtcctg
421  clarratttgc tggccccccg gtagatgttt cccccggag ccccatcgc acacccacgt
481  cllccccca cctctcacc cccclcatgg tggctgctgt tcccatcag gacccgntgg
541  acaaccgggt ggcannggcg agcagagaga gcccggcgag aggggtgagc cggagcctcc
601  tggatccccc gclccctccg gctctcttga ccccgaggla caccaccaaa tggggcanncc
661  ccccaatttt aagagctcnc ggcgcccaag cagacacact gaggctcccy lccgatttg
721  laacatgttt ttctctccl cclagatgcc aaccggtgac aclygclcca aagagctcg
781  tggggctong gclcccccct gtagtaccgg ggggagagcl acagagtgag gaggagagcg
841  cgtggggct gactgtgtc tagcgglll cctctctccc tctctctct gclctcttga
901  lccctcccc aagagagcnc gggattcccc gtagctgag acgclgclgg accgccccgc
961  actaalglga gtctaaaaac gttclgggat tccccccacc tgggglltgg gcgtgcttc
1021  cccgccttgc gtgttggagg gggcctgttt tccctgacac gacagctgag gtttccctc
1081  ttggtctctc gclgttggct ttgggggccc lcccaatgag agagagggac ttclclagag
1141  gcaagagctc cccannggcg agcagtgagg tgggggtgcl aagatagag tggctgctc
1201  lgggttgac caacgtgtcl lccctccct ttagggtag cccggccccc cggagccccc
1261  cggctctgtc ggcagagagc accccagggg gcllctgagc gacgcccggc ccccgggcgc
1321  lgcaggtgac cccggcctcc agggccccgc cggccccccc ggcggagagc acgagccccc
1381  cggagagggc cccgagatga gattctggg gclctctccc ctccgtgac cccclgctc
1441  cglggggcgg ttgtcttag ttgcltccc cctctgctc cctgagggg lccgagagc
1501  cccccggcc clcannggct ggcagagagc gttgglltgg tgggtctccc agggagggct
1561  ggtgggggag gcttccccgg agtggcgggc cctcagatga gggggcgtc ctcatatgag
1621  tgcactgagt cctclggggt gcagagatgt gggggccgag atgctctggg gcccclctca

```

批注: The nucleotide sequence (part sequence of the genome)

1681 AAAATGCGA GCGCTTAT RT GCGCGCGGGG AACACGTTT GCGGTGGGT TAAATTTTT
1741 CGGAGTCTT TTTTCCCHC CAGGAGCGG TGGTGCAGG CTTAAGCGG GTCGAGGAAA
1801 ACCGTCGCT GTTATTGTT AGCGAGGCT CTTGGAGCGA TAAAGGAGG GCGAGGCGA
1861 GCGGTGGGT CTTCCGCT CATGCGCT TTTCCGCT TCGAGGAGG AGCTGGAGG
1921 CAGGAGGCT CTTGCTCTC AATGAGCGG GTTCCCGCG GCGCGGTGGG CCGCGCTGGG
1981 CTGACAGCT CTTGCTGCGA ACCGAGGCG GCGGAGGCG AAAACCGCG GAGTACGCG
2041 GCGGCGGCT ATCAGGAGG CCGTGGGCG GCTGAGCTC CAGGCTGCG CTTGCTCTT
2101 TTTCCGCGA CTGAGGAGG ACTTTGCGA GCTGCTCTC CCGCGGTGGT GTAGCTCTT
2161 CTTTCTCTT TGGGCGGCT CTGAGCTCTC GGTGCGCGG GCGAGGAGT GCGAGGCTG
2221 CATGAGGCT GTTCTGCTC AGCTGCGCG TTTGCGGCT CCGTCCGCT GCGAGGAGT
2281 GTGAGGCTT GTGCGGCTC AGCTGCTCTC TTTGCTCTC GCGAGGAGT CATGAGGCT
2341 CCGGCGGCT AATGCTCTC GTTCTCTC GCGCGCTGG GCGCGCGGCT CTTGCTCTT
2401 CCGTGGAGG AGCGGCGGCT AGGAGGAGG CCGTGGGCT TGGGAGGAG GTTATGAGT
2461 TGGGCTCTC TTTGCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2521 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2581 CCGAGTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2641 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2701 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2761 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2821 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2881 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
2941 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3001 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3061 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3121 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3181 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3241 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3301 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3361 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3421 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3481 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3541 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3601 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3661 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3721 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3781 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3841 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3901 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
3961 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
4021 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
4081 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
4141 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
4201 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC
4261 GTTCTCTC GTTCTCTC AGTGTCTC GGTGCGGCT CAGGAGTCTC TCGTCTCTC

4321 gggagagagc gggggaccac caggagagaa gacagagagc agctctggtt tggagagacc
4381 atcaacggcg gttccacgt ggggagccc cagggtgctt tggagagac gttccacgt
4441 gggatgctct tctggggtc atgagagagc gtttaataa gtttggagc ggtgattctc
4501 aagggtgagc tttgagtcag ttgagtcga caaatcagag gggagagagc ggtgagagc
4561 actgagtcct ctgaggttgc tggggcaga aagagagagc atgagagagc ggtgagagc
4621 cctacagccc ctgagagccc tacagagagc ttgagagagc agttctgccc caggagctgc
4681 ttgggagagc aacagagctg tggagagagc ctgagagagc gttgagagc aggtctgagc
4741 acattgggtt tgggagagc ctgagagagc gttgagagc gggagagagc aggtctgagc
4801 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
4861 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
4921 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
4981 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5041 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5101 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5161 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5221 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5281 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5341 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5401 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc
5461 tgggagagc gggagagagc aggtctgagc gttgagagc gggagagagc aggtctgagc

Attachement 4: Search record for AF452711, version AF452711.2 GI:31340541

LOCUS AF452711 12003 bp DNA linear VRT 08-FEB-2006

DEFINITION Gallus gallus alpha 1 type II procollagen (COL2A1) gene, exons 8 through 52 and partial cds.

ACCESSION AF452711

VERSION AF452711.2 GI:31340541

KEYWORDS .

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 12003)
AUTHORS Xi, C., Liu, N., Wang, F., Guo, S., Sun, Y., Yang, F. and Xi, Y.
TITLE Molecular cloning, characterization and localization of chicken type II procollagen gene
JOURNAL Gene 366 (1), 67-76 (2006)
PubMed 16297573

REFERENCE 2 (bases 1 to 12003)
AUTHORS Caixia, X., Yongzhi, X., Siqi, G. and Yuying, S.
TITLE Gallus gallus alpha 1 type IIA collagen precursor (COL2A1)
JOURNAL Unpublished

REFERENCE 3 (bases 1 to 12003)
AUTHORS Caixia, X. and Yongzhi, X.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2001) Immunology, Beitaipinggu Hospital, Beitaipinggu, Beijing 100039, China

REFERENCE 4 (bases 1 to 12003)
AUTHORS Caixia, X. and Yongzhi, X.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-2003) Immunology, Beitaipinggu Hospital, Beitaipinggu, Beijing 100039, China

REMARK Sequence update by submitter

COMMENT On Jun 3, 2003 this sequence version replaced gi:28190290.

FEATURES
Location/Qualifiers
source 1..12003
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
gene <1..>12003
/gene="COL2A1"
mRNA join(<1..54, 179..232, 340..393, 494..547, 646..699, 786..830, 941..994, 1086..1130, 1244..1297, 1387..1485,

批注: The nucleotide serial number of the genome

批注: The sequence of other vertebrate

批注: This is the sequence submitted for the last time after the paper was published.

批注: This is the date when the sequence was submitted for the first time; but the nucleotide sequence submitted this time was not disclosed, and no relevant sequence can be found in Genbank.

CDS

1684..1728,2096..2194,3080..3133,3542..3649,3738..3791,
3883..3981,4664..4717,4800..4898,4981..5034,5159..5212,
5490..5543,5623..5676,5791..5835,6098..6196,6289..6396,
6477..6530,6612..6665,6748..6801,6896..6949,7039..7146,
7255..7308,7422..7475,7744..7905,7998..8105,8415..8522,
8683..8736,8834..8941,9210..9263,9373..9480,9615..9668,
9837..9944,10024..10312,10720..10907,11502..11744,
11857..>12003)
/gene="COL2A1"
/product="alpha 1 type II procollagen"
Join(<1..54,179..232,340..393,494..547,646..699,
786..830,941..994,1086..1130,1244..1297,1387..1485,
1684..1728,2096..2194,3080..3133,3542..3649,3738..3791,
3883..3981,4664..4717,4800..4898,4981..5034,5159..5212,
5490..5543,5623..5676,5791..5835,6098..6196,6289..6396,
6477..6530,6612..6665,6748..6801,6896..6949,7039..7146,
7255..7308,7422..7475,7744..7905,7998..8105,8415..8522,
8683..8736,8834..8941,9210..9263,9373..9480,9615..9668,
9837..9944,10024..10312,10720..10907,11502..11744,
11857..12003)
/gene="COL2A1"
/codon_start=1
/product="alpha 1 type II procollagen"
/protein_id="AA033039.2"
/db_xref="GI:31340542"
/translation="GPQGFQGNPGEPEGAAGPMGPRGPPGPGKPGDDGCTGKPKG
SGERGPPGPGQARGFPCTPGLPGVKGHRYPLDGAKEAGAPCAKGESGSPGENGSP
GPMGPRLPGERGRPGSGAAGARGNDGLPCPAGPPGPVGPAGAGPFGPAGSGKEAG
PTGARGPEGAQQPRGESGTPGSPGAGAPGNPTDGIKAKGSACAPGIAGAPGFPPI
RGPPGPQGATGPLGPKGQTGEFGIAGFKGEQCPKGETGPAGPQGAPGPAGEEGKRGAR
GEPGAAGPVGPPIGERGAI'GNRGFP'GQDGLAGPKGAPGERGPAGLAGPKGATGDPGRPG
EPCLPGARGLTGRPGDAGPQGVPTGAPGEDGRPPGPGQARGQPGVMGFP'GPKGA
NGEPKGAGEKGLPGAJ'GLRGLPGKDGETGAACPPGPACPVGERGEQGA'G'PSGFQGLP
GPPGPPGESGKPGDQCVFGEAGAI'GLVGP'RGERGFPGERCSPCAQCLQGPRGLPI'GT'PG
TDGPKCATGPAGPNGAQGP'GLQGM'GERGAAGTAGPKDRCQDVGEKGP'EGAPGKDGA
RGLTGP'IGPI'GPAGPNGEKGTSGPPGPGSGAAGARGAPGERGEPCAPGAGFACPPCAD
GQ'FGAKGEQGEPOQKGDAGAPGPGGP'SGAI'G'QGPTGVTGPKGARGAQGPPGATGFI'G
AAGRVGP'PG'NGNP'GPPGPPGCSAGKDGPKGVRGDAGFP'GRAGDPGLQGPAGPPGEKGE
PGEDGPAGPDGPPGQLAGQRGI'VGLPGQRGERGFPCLPGPSGEI'GKQGA'GSAGDR
GPPGPVGP'GLTGPAGEPGREGNPGADGPPCRDGAAGVKGDRGETGFPVGPAGAPGAPG
APGPVGP'TGKQGDRETTGAQGPMPGSPGAGARGMP'G'PQGP'RGDKGETGEAGERGLKGH
KCF'TGLQGLPGPPGPGSGDQGAAGPAGPSGPRGPPGPVGP'GSKDGSNGMPCFI'G'P'GPR
GRSGEPGAPGPPGNP'GPPGPPGPGTGTIDMSAFAGLQGT'EKGPDI'IRYMGADEAAAGGL
RQHDEYVDATLKSNNQIESIRSPEGSKKNPARTCRD1KLCHPEWKSQDYWTDPNQGC

TI.DA|KVFCNMETGETCVYPTPSSTPRKNWTSKTKDKKIVWPAET|NGGFHFSYGDE
NLS|PNTASIQMTPI.RLLSTEGSQNVTYHCKNSTAYMDEETGNLKKAILIQSNDVEIR
AEGNSRFTYSVLEDGCTKHTGKWGKTVIEYRSQKTSRLPIVDIAPMDIGGADQEFQVD
TGPVCPL~

exon <1. . 54
/gene="COL2A1"
/number=8

exon 179. . 232
/gene="COL2A1"
/number=9

exon 340. . 393
/gene="COL2A1"
/number=10

exon 494. . 547
/gene="COL2A1"
/number=11

exon 646. . 699
/gene="COL2A1"
/number=12

exon 786. . 830
/gene="COL2A1"
/number=13

exon 941. . 994
/gene="COL2A1"
/number=14

exon 1086. . 1130
/gene="COL2A1"
/number=15

exon 1244. . 1297
/gene="COL2A1"
/number=16

exon 1387. . 1485
/gene="COL2A1"
/number=17

exon 1684. . 1728
/gene="COL2A1"
/number=18

exon 2096. . 2194
/gene="COL2A1"
/number=19

exon 3080. . 3133
/gene="COL2A1"
/number=20

exon 3542. . 3649

exon /gene="COL2A1"
/number=21
3738..3791
/gene="COL2A1"
/number=22
exon 3883..3981
/gene="COL2A1"
/number=23
exon 4664..4717
/gene="COL2A1"
/number=24
exon 4800..4898
/gene="COL2A1"
/number=25
exon 4981..5034
/gene="COL2A1"
/number=26
exon 5159..5212
/gene="COL2A1"
/number=27
exon 5490..5543
/gene="COL2A1"
/number=28
exon 5623..5676
/gene="COL2A1"
/number=29
exon 5791..5835
/gene="COL2A1"
/number=30
exon 6098..6196
/gene="COL2A1"
/number=31
exon 6289..6396
/gene="COL2A1"
/number=32
exon 6477..6530
/gene="COL2A1"
/number=33
exon 6612..6665
/gene="COL2A1"
/number=34
exon 6748..6801
/gene="COL2A1"
/number=35

<u>exon</u>	6896..6949 /gene="COL2A1" /number=36
<u>exon</u>	7039..7146 /gene="COL2A1" /number=37
<u>exon</u>	7255..7308 /gene="COL2A1" /number=38
<u>exon</u>	7422..7475 /gene="COL2A1" /number=39
<u>exon</u>	7744..7905 /gene="COL2A1" /number=40
<u>exon</u>	7998..8105 /gene="COL2A1" /number=41
<u>exon</u>	8415..8522 /gene="COL2A1" /number=42
<u>exon</u>	8683..8736 /gene="COL2A1" /number=43
<u>exon</u>	8834..8941 /gene="COL2A1" /number=44
<u>exon</u>	9210..9263 /gene="COL2A1" /number=45
<u>exon</u>	9373..9480 /gene="COL2A1" /number=46
<u>exon</u>	9615..9668 /gene="COL2A1" /number=47
<u>exon</u>	9837..9944 /gene="COL2A1" /number=48
<u>exon</u>	10024..10312 /gene="COL2A1" /number=49
<u>exon</u>	10720..10907 /gene="COL2A1"

/number=50
cypu 11502..11741
 /reno="COL2A1"
 /number=51
bxon 11857..>12003
 /reno="COL2A1"
 /number=52

ORIGIN

```
1  ggtccccagg gatttcnngg caacccccagt gagccccggc gacccaggcg tgcgtgagt
61  gtcccatcct tgcctcact ctgcgccctt ttgcatccc cclcnactt gggacgtgtc
121  tccgtccctc ctaccnuc cccccccac ctacnctcgt tctctttgt cccgtlugg
181  tccatgggtl ccccgaggac ctccgggccc acctgggaaa cccggtgncg atgtgactcc
241  cccaglmga gcratctra gcnccggcgg ctcatgcaa tgnlgcccc cttcagacc
301  ctgcttctc tccttlnaa ccccatntt tccclgcggg rtgagacagg cmmccccggc
361  uatctgagt aacgtggccc ccccgkcccc caggtagggc agcuctgggc tgaacgctac
421  cccccnlg clyccccaa cctttttcc cclcccccc tcaagtcaa cctnccccu
481  lllgtctcca caggagcgtc glygttccc tgggactcct ggttccccg gnglgagag
541  cncggagtg agtcccccc acccncucc gccccccat cacttatatc cncgglygc
601  ccccatana gllgllcgt cctctctcc ttcnccnna ncngggcta ccccgttt
661  gatggtgcca naggagaggg gggggctcct ggagccnna tngtggggg ggcagccct
721  ccatggcgt gkcatgtta ccccaacct gggctcctc lclctgacc cttttccca
781  tacaggtat nctlgllca ccgggtgaga acgactccc cggccnctg gnglygcyg
841  gggttggca tgggggaaa tgggggagc gnllyggg cggccccat aacctatgc
901  ctgagccna cccnccctg tccccctcc acccccgag ggcacccgtg ggclycccc
961  agagcgggn cgtcccgccc cclccgccc ccccttact rtgacagtgg cctntggcn
1021  cngggctgcy tglgggggg gacggggtt ctgtgacccc cccnctgct gcgctctcc
1081  cccaggttac tctgagcatt gacggglna nlygcccgc tggccccct gtaagtata
1141  ggtgtgttg acccncucc lcnctllct nctgaacccc rtggggttc tccgggagcc
1201  cccccagct gggcagccc cccnctctt cctgllccc cngggggccc tggccctgc
1261  gggngcccc ggttccccg gggccccgg ttcaaggtt aggggtgac agaggagct
1321  gggggggact ggggggggg aaactgagg gcagtgccc gtctcncnc ncttctctc
1381  ccccgaggtt gggcagccc cctggggn gggggllccc ggggtgccc aggacccgc
1441  ggggagtcg gggcccccy cclccccgc cccgctggg caccgttgg tctccnct
1501  ggggagtcg cctggggtt cccatctcg tggatcccc tgggctggg gngctgltg
1561  gllgggggca cgtgggggt agtggggtg gggggggatc tgggtggcg gggggllatc
1621  ctgtctgtg ggtggggga nctggggcag ggggtggag atgtctcct gttgttttt
1681  tgggttacc cgggagctt lgggllccc ggtggcagg gctcagcgt ggttttggc
1741  cgggtgggg ggggggggg ggtccgaaa cctttgggt ggggtggly gllcnctlyg
1801  ggggtgltg gnglctaat gatctgggg ggggggggt ggggtgltc ggggggtac
1861  ggggggtgc tgggggctc ggggtgctg ttggggggg ggggtgctg ggtgtgtta
1921  lglglgtgg atcgggggt cactgtgga tglggggg cggggggg tggggggga
1981  tgggggggt cgggggggt lgtgggggt cggggggga tgggggggt cgggggggt
2041  tggggggga tgggggggt cgtgggggt cggggggga tgggggggt cgggggggt
2101  cggggggtt ggggggggt cgggggggt cgggggggt cgggggggt cgggggggt
```

2161 tgcacccggg cccctgrrac cccaaaggucg gacrrtaaga gaccccccga gaccccccacg
2221 tgggggcccc grrtatrrcc cggccactg cccarrtggc cggcccccac tggaggttgc
2281 cccgcccact ttgcccggg ccccgcccac cglgcccggg kacrcgcc clycugrrcc
2341 cccmlllgrr gacactctgt tcccclccc aaagtcccc atggtcactc ccaggagacc
2401 clcttctrrt cccmwwggg gccactrrrr atgggttggg tggtrrrrtt ggggtgggt
2461 gggggttggg tgaratrra tggggttggg acggttrrr rrrtgggtt ggcactrrrr
2521 tgggtatagg gacggggggg aggatrrggg aagggttggg kacatrrgt ggggtgggt
2581 gggggttggg gllkrrrra tggggttggg gggatrrrra aggggttggg gclgggggt
2641 gggggggg ttaaggggg gggggttggg gttaggggtt ggggttggg lrrrrrrrra
2701 gctgggggtt gllggrrrra rrratrrrrg ggggttggg gggggggg atgggggtt
2761 crrrrrrrra tgggttggg atgggttggg rrrrrrrrrt gttgggttggg ggggggtt
2821 ggggggtt acarrttrrr ggggttggg gllggrrrra aggggttggg tgggggtt
2881 rrrrrrrrra ggggttggg ggggttggg rrrrrrrrra atgggttggg ggggggtt
2941 ggggggtt ggggttggg tgggttggg tgggttggg atgggttggg ggggggtt
3001 ggggggtt cgggttggg rrrrrrrrra ggggttggg ggggttggg cgggttggg
3061 cgggttggg tgggttggg ggggttggg ggggttggg tgggttggg ggggttggg
3121 ggggttggg ggggttggg cgggttggg cgggttggg cgggttggg cgggttggg
3181 cgggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3241 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3301 ggggttggg cgggttggg cgggttggg ggggttggg ggggttggg ggggttggg
3361 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3421 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3481 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3541 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3601 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3661 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3721 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3781 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3841 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3901 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
3961 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4021 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4081 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4141 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4201 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4261 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4321 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4381 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4441 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4501 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4561 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4621 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4681 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg
4741 ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg ggggttggg

1801 klyctecteg caurghnlyuc cccccccccc cccccggacc tcagggllyct cgttarrcagc
1861 ctgglyglyat aatgtttccc ggtcccaaaq pggctnnlyt aatgtttgac cattttctgc
1921 cccnnatggg aactatgung gtcagccagc ctgtgugccc atcccgcttc ctctctgtag
1981 kktkagcctg gnnnngctga aagagnggga ctgcccagcc cccnngggct acatgtgagc
5011 tcccttgccc actctgaccc anlnnntca ccttggtggg ttcccccgc actgltccct
5101 glnatgtgat aagghnlyuc agccacagag clyngtggc aatgtttctc ctctgagag
5161 tctkccctggc nngghnlyuc aagcagagc tgcagccccc cctgggccc ctgtatgt
5221 cctttlllll cctcttgagg acagatgttt gnatatgagc gctggggggg glnnccccca
5281 aatctggggy gaagaccagc cngghnlyuc tggggtatgc ggtgggggaa aagccccggl
5341 atgtgnglyl gggatgttgg tglcccgccc ccccatccc ggtgtgattt aatgtttgtc
5401 cccnaccctt aatgtgnglyl gggatgttgc aggggtgagc cccagagagc gngghnlyuc
5461 cagatgtgic actgtatgc tccclllagg gctctgtagg tngghnlyuc aagcagagag
5521 cccccctcc tccgggllc aatgtatgtg ggggagagc aagagaggtt ggtgtgagc
5581 aatgggagag cccccctac accatctt. gtccccaac aggggtgag cngghnlyuc
5641 aatccccg gngghnlyuc cccnngggg gggatgttgc aggggtgag cccatgttgc
5701 cngghnlyuc cttatgtgt gngghnlyuc ccccccgat gngghnlyuc cccccagtc
5761 tgaacccctg ctctctgc ctctctgag gngghnlyuc aagagagagc tggccccggt
5821 ctgtgtgag cagatgtgag llllyctgt ccccccgat cctctgtgc cactgagag
5881 tttgggggt gngghnlyuc cagatgttgc ccccccgat cagagagagc gngghnlyuc
5941 ngghnlyuc actgtatgc cngghnlyuc tccagatgc cccnngggg accgtgag
6001 aatgtatgc ctctctgc aagagagag cctgtgtg. cctctatgc actgtgag
6061 gctctctgc actgtatgc cctctgc cccatgttgc aagagagat tccccggl
6121 acgagagat cccgtgagc ngghnlyuc aatgtatgc ggtgtgagc gngghnlyuc
6181 aatgtatgc cccnngggg aagagagag atgtgtgagc agtgtgagc accatgtgag
6241 cgtgtgagc tccagagagc ngghnlyuc tccatgttgc actgtgtg. tgggagagc
6301 cngghnlyuc cccagagatgc cagagatgc cngghnlyuc agggatgc cagagagag
6361 aagagagatgc actgtatgc cccnngggg cccagagatgc aatgtatgc cccnngggg
6421 tccgtatgc cngghnlyuc cagagatgc cagagatgc cngghnlyuc cccnngggg
6481 atgtgtgag gngghnlyuc gngghnlyuc cagagagagc tgggagagc gngghnlyuc
6541 cngghnlyuc gngghnlyuc cagagagatgc cngghnlyuc cngghnlyuc actgtatgc
6601 ctctccata ggtgtgagc gngghnlyuc gtcctctgc cctgtatgc cccnngggg
6661 agnnggtgag ngghnlyuc cagagagagc cttatgagc cngghnlyuc cccnngggg
6721 tccacctgc cctcttct tttccagggg gngghnlyuc ctctgtatgc actgtatgc
6781 gngghnlyuc kktgtgagc agtgtatgc atgtatgc cccgtgtgc cccatagct
6841 tccacctgag cccnngggg ccccttgc accctgagc actgtatgc actgtatgc
6901 ggtgtgagc cccgtgagc ngghnlyuc tggatgtgc aagagagagc tnggtgtgc
6961 acccagagc cccnngggg cccagagatgc tccacagagc actgtatgc cctgtatgc
7021 kktgtatgc cccataggg ggtgtgagc cagagagatgc cagagagagc gngghnlyuc
7081 cccagagagc ngghnlyuc gngghnlyuc aatgtatgc ggtgtgagc ggtgtgagc
7141 cccagagatgc aagagagagc gngghnlyuc cccagatgc aagagagagc gngghnlyuc
7201 agggagagc actgtatgc tggatgtgc ngghnlyuc ctctctgc cagagagagc
7261 accagatgc atgtgagagc ngghnlyuc aagagagagc ggtgtgagc ggtgtgagc
7321 aatgtatgc cngghnlyuc aagagagagc ggtgtgagc ggtgtgagc ggtgtgagc
7381 tccctctgc ctctctgc actgtatgc tccctctgc gngghnlyuc gngghnlyuc

7441 GAGCTGCCRR CCGTGLUUUU CCGCCCCCCC CTANTGLGUG TCTKKKKKCG TTCTGGGULL
7501 GCCCOCUUUL KUUUUTTTGGG CACTGCTTCC CCGGCTGCGG TGTTRRRRRR GGCULGTTT
7561 CCGTGCACAG ACACATGGGG TTTTCCCTCT TACTCTCTCG ATGTLGGCTT TTGGGKCCAT
7621 TCCCATRTTA AGAAGGNNL TTTCTAUGGA CAAGAGCTCC CCGUAGUGCA GCATRRRRAT
7681 GCHHRTATA AAGATGGUUL GUCUGCTCT KATTGACCC ANCGCTGCTT CCTTCCCTT
7741 TARRKTAACC CAGGCCCCCC CCGCCCCCCC KACTCTGCTG GCAAGGACUG CCCCBAARRAT
7801 ATTGGLGGG AGCCCCGGCC CCCCRRCCAT RONGGTGACC CCGGCTTCCA AGRCCCCGCC
7861 GGGCCCCCGG GCGAGGAGRR CKAACCCGGC GAGGAGGGCC CCGCGGLGUG KATTCTRRRR
7921 GTCTCTCCG TCCGLGCC CCCLKCTAC GTGGTGGCT TGTCTTNGT CTGULLTCCC
7981 CCTCTGCTG CCGGUGGUL CCGGACGGCC CCCCCRRCC TCAAGGCTTG GCGGGNUGG
8041 GTGGTATTGL UGTCTCCCA KACACAGTG GTGAGAGAGG ATTCCCGGA CTGCCUUUUG
8101 CATCAGTGAG TGGTGGGLC LCATLLGGU GCACTGATC CTATRRRRR CARAGATGT
8161 GUGGCCCCGA TCTCTRRR CCACTCTAG GGTGCGCG CCGTTTGGTG CAGCCCCGGG
8221 ACACGGTTTG CUGTGLUULL KKKTTTTTC KRGCTCCTT TTTCCCAAC AGGAGCGCT
8281 KRTGCAAGG TTAGGCGGG GCGGGNNNN CCACTGGLG LLLLLGLLY CAGUUUUULC
8341 TGGGNGCUL NNNNNNUGG UUUUUUUUG CUGLGGULC TCTCCACTC ATGCACCTC
8401 TTTCCATCTT TCAARRAGA CTRRAAAGC AARRARACC TACTCTGCG KRTACCCAR
8461 KCCCCCCGH CCCCRTKRC CCCCCTKKC TACHHHTCC TCTKKKAKA CCKKKKCKH
8521 AGGTANGCU AACCCCHNG CATCHNGG GACCGGGCH TCCCHNCCC CATGGCACG
8581 CTCAGTCCC AGAGTCCCC KATCTTTTT TCTCCACAC TGAARRARA CTTTACACA
8641 ATCTGTCTC ACCCGRTTG TAACATCCC TTTCTCCT AARRCAACC TGTGCLGHC
8701 GGLNCCCHG GNGGGNTGG CGNGCTGAC GTGAARRTA KCTTCCATG CACTCCCCAT
8761 TGGNCTGCG CATCCCCR CCAAAAGCL KKKKLLLLG CACHGATCG ACCTCTGT
8821 TGLTGTGCG NGGGTGTG GTGRTGAGC CRRCCCTGT TGTCTCCC GLTCTCLGG
8881 ARCCCCRR CCCCCGGCC CTGTGTG CCLGGNNNN CAGGAGACA KARRCGAGC
8941 GGTGAGTGT KACACAARR TTTARRTTT AUUULCTCT TATGGCTGA AATGTACAG
9001 KRTTCCCCC AARTTTTGT CCTLGCACH GLTGNGTG CATTTAARA TGTGTGUGG
9061 CCHGNGAGG TGTGATTGT CACTTTGCC CGBULCLUG GTGCGGACA TARRCTRR
9121 CCGGACCCC CCGNNATCC AATTCATG CTTGAGGLY GLTCTCTGG TCGTGGCAC
9181 CLTGTGTG CTTTTTTT TCCCTGAGG GLGCACUUG GGCATRRRT CCTCTGATC
9241 CCGCTGGGC TCGUUGULY CCGGTGATG RTCTGATG CATCGUACH TCCCHGTAC
9301 AAGACTRRR GTCTGCTG CCGUUGUGG GTGTGCCAC CTGAGCCGA CACGCGCTG
9361 TCCCCCTTT AKKKTCCCC AGGUVCTGT GGTGACAAAG RTARRACRR AUUUUCLGGA
9421 GUGUGUGGG TGNNGGACA CCGCAGCTT ACCKKTCLY AUUGLCTGC CCGACACCC
9481 GTAAGTTGT TARRRRAGA CTGAGCCCC CCCCCGTAC GATGCGCTC CTTTRRRGT
9541 TGLTGGGCU CCGAGGTGT GTCTRRCCA AARTGTGAC CGCAGGCTG LGCCHCCC
9601 GULCTCCLC LAGGGGCGG TCTRRAGCC AARTACTAC CCGTCCGCT GGLCCCTCG
9661 KCCCCAGL UHLCCLGHC GUTGTTT GGGTGGTGG AAGGGGAGG AGCAGCAGT
9721 GCTCCCTRR GCACCTGAG CCTGTGTC CTCTGTCTG CLCHLNGCU CCHTGCCTT
9781 CCGTGCCTG AGCCCCGCA ATGCCILHC CCCCCGTLT LGGGGTCTC TCTAGRRTC
9841 CCGCTGTC CACTRRCCC TCTRRCAAG ACSTCTTA CCGCATGCC GGCCHCTG
9901 GLCTCCCGG TCCCGTGGG CGGAGTGTG AACCCGACC TCGGTGAGT CCTGTGAGG
9961 KRRRRCAARR AATRRRLC AGCLGCHG GNGCCCTC AGCTACCTT CTTCTCCCA
10021 LUGGLCTC ATGGNNNN CGLCTCTCC KRTCTCTG RCCCCCRR CACCGCATC

10081 gaculglclg cttttgctgg actgagtcag acggagaggg gccccgaccc catccgtac
10111 atgggggchg acgagggagg cggagggtg cggcgucacg acgtgaggtt ggucggcucc
10201 ctcaaatccc tcnucalca gattgagagc atccgcggcc cggagggctc caagaagunc
10261 cctgccagga cctgcggcga ctcnuaactc tggcatcccg agtggungug cggtaagagc
10321 tcccgglgcc tctcccgtaa tccctcttc cccacggagc agcatccccc ggcctcctgc
10381 accgcctgc gglcugklk gattgtagga angattcctt gtcnuaagga gctctgagcg
10441 ctgagctgga ctgcggggg agglgggga gtcgctgccc ccatagggtt tgggggucgt
10501 tggagatgta gcacttggga gcglggccta gtggggatga ggcagcagtt ggaaccatct
10561 tggaggtctt atccgclclt atggctctg tctctctgc ggtgtgctg gtggctgctg
10621 gtagccattt agacttggcg atcttgggg tcllllccga tctaacgac tcttagacct
10681 ccccnacccc atgagcgctg ttgctctcc cccctgagc agattactgg attgacccga
10741 accagggctg cacttggac gcaatcagg tttctgcau culggagaca ggcgagacct
10801 gctctacccc gacacccgca agcclcccc gggagagctg gtagccagc aagacgaaag
10861 acnaggagca cgtctggtt gcaagagaca tcaacggcg tttccnctg ggtgtcccc
10921 aggtgtctt ggaaggatg atccnctg gglgclctt ctggcgta tctgaggg
10981 tlllualgga gllalagagc gtagcttga aggtgtaggt ttgggtcgt tngctccc
11041 aatcnaagg gnaaggntg gtaggggca ctggtctcc tgggtllgt lggcccgga
11101 aggtgagga tggggggg cclcgggcc clacggccc ttagggcct acagcgct
11161 taggaaaaa gttctcccc gtagctggt tgggacaga acagccctgt gtagccgag
11221 ctgagggagc attgagaca gctctcga cttgggtt ggtcnggtc ctgggttng
11281 ttaggtgagc ggggcnacn ggcctgggt ggtgttng ggccttca agcnnct
11341 tccatgttc ttagctctg agggacctt ccaatccaa cactctgat tttttctc
11401 gcatcttgg agcctgaggt acggaggtc tcccaaaag ctctgagag taaggtgct
11461 ataatgccc caggtcttca ctctcctt ctccctca gttcagctc gtagctgga
11521 accclgcccc cnaacccgc agcclccga gaccllctt ggccllctt tcccccggg
11581 gctcccgaa cgtcacctac cactgcaaga acagcatgc ctacatgag gaggagcgg
11641 gcaacctga gnaagccatc ctcatccag gacccagca cgtgggagc agnccggg
11701 gcaacagcag gttcacctac agcgtcttg aggagggctc cagglaggt lgcgggctg
11761 ctgcnaggga aggtgcnag lgggggggg gtagctgag ctggggggt gaggccgag
11821 cagctgagc catccctgc ctcttccct ccccgagac acactggca atggggctag
11881 acggtgagc agtaccgct gcaagagacc tgggctgc ccatgtaga tattcacct
11941 atggacattc gggagccga tcagggttt ggcgtgga tggggcggt ctgctctctg
12001 tan